IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received January 12, 2019, accepted January 30, 2019, date of publication February 8, 2019, date of current version March 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2898265

EdgeCare: Leveraging Edge Computing for Collaborative Data Management in Mobile Healthcare Systems

XIAOHUAN LI[1,2], (Member, IEEE), XUMIN HUANG[3], CHUNHAI LI[2], RONG YU[3], (Member, IEEE), AND LEI SHU[4,5], (Senior Member, IEEE)

[1]School of Electronic and Information Engineering, Beihang University, Beijing 100083, China
[2]School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China
[3]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[4]College of Engineering, Nanjing Agricultural University, Nanjing 210095, China
[5]School of Engineering, University of Lincoln, Lincoln LN67TS, U.K.

Corresponding author: Rong Yu (yurong@ieee.org)

This work was supported in part by the Programs of NSFC under Grant 61422201, Grant 61370159, and Grant 61762030, in part by the Science and Technology Program of Guangdong Province under Grant 2015B010129001, in part by the Guangxi Natural Science Foundation under Grant 2016GXNSFGA380002 and Grant 2018JJD170022, and in part by the Key Science and Technology Project of Guangxi under Grant AA17204009 and Grant AA18242021.

ABSTRACT With the wide application of mobile healthcare systems, the total amount of healthcare data is ever increasing rapidly as users interact with healthcare service providers frequently. This leads to a challenging task to manage healthcare data. Existing work mainly pay attention to centralized and blockchain-based mechanisms. But they cannot adapt to the increasing amount of global healthcare data and suffer from complex application challenges, respectively. Decentralized and collaborative data management assisted by edge computing exhibits major advantages in improving overall system performance. We present a secure and efficient data management system named as EdgeCare for mobile healthcare systems. Local authorities are established to schedule edge servers for processing healthcare data and facilitating data trading. A hierarchical architecture with collaboration is designed for feasible implementation of EdgeCare. After that, we investigate secure data uploading and sharing in the system. We use an electronic medical record to show how healthcare data is processed with security considerations. We also conduct the Stackelberg game-based optimization algorithm to approach the optimal incentive mechanism for a data collector and users in the fair decentralized data trading. The numerical results with security analysis are provided to demonstrate that EdgeCare offers effective solutions to protect healthcare data, and support efficient data trading.

INDEX TERMS Public healthcare, collaborative work, internet of things, distributed management.

I. INTRODUCTION

Recently, mobile healthcare systems have been applied widely because of the advances in biotechnologies, information technologies and software engineering. Due to pervasive e-health services, billions of personal healthcare data records are generated annually [1]. Meanwhile, users interact with healthcare service providers frequently in current applications, e.g., mobile telemedicine, personalized medicine and emergency response [2]. Towards secure interaction, a tremendous amount of healthcare data should be collected and processed well. Besides, healthcare data is a valuable commodity for public health agencies, academic researchers and pharmaceutical companies to promote service provision, including disease prevention and control, healthcare-related researches and drug development, via big data analytics [3]. Here, a typical commercial case of data mining in mobile healthcare is about Google. To develop a diagnostic app, the subsidiary Deep Mind collects and analyses healthcare data records regarding to 1.6 million patients in hospitals of London, UK [4]. Before operating effective data mining, third parties need to gain authorized access to healthcare

The associate editor coordinating the review of this manuscript and approving it for publication was Joel Rodrigues.

data. Data trading is necessary to negotiate with users as data owners. As a consequence, there exist critical issues to manage numerous healthcare data in a secure and efficient manner.

To tackle the problems, some of existing work pay attention to centralized data management with cloud computing in mobile healthcare systems. Centralized data management leads to high-latency response and excessive workloads when the central authority continuously processes a lot of healthcare data over time [5]. The central authority also easily suffers from typical single point failure, DDoS attakcs and remote hijacking attacks. So a majority of work have begun to focus on decentralized data management for performance improvement in recent years. In particular, they leveraged the blockchain technology which is a peer-to-peer distributed ledger to prevent any repudiation in the trustless environment and achieve totally self-organized and transparent management by depending on a set of consensus nodes. They aimed to facilitate authentic, verifiable, and traceable data management in healthcare domain.

But most of the blockchain-based healthcare systems are integrated with cloud computing. Healthcare data is still mainly stored in the remote cloud [6], [7]. This cannot support prompt data transfer and real-time data access. Besides, we consider that there exist crucial basic problems, e.g., node selection, consensus mechanism and blockchain interoperability, needed to be strictly solved for fully considerate deployment of the blockchain-based healthcare systems. Due to initial complexity and application challenges, current work mostly pay attention to future outlook [8], architecture design [9] and prototype implementation [10] in the systems. Moreover, practical data trading is seldom investigated to build a healthcare data market in the researches. Thus, we are motivated to introduce edge computing in decentralized data management and simultaneously consider data trading for the improvement of overall system performance.

Decentralized and collaborative data management assisted by edge computing shows great advantages in the facilitation of mobile healthcare systems. Edge computing has been envisioned as a crucial enabling technology to enhance computing environment by deploying ubiquitous communications and proximal cloud computing capability, and to serve users in vicinity. The core concept, applications and key technologies about edge computing are presented in the white paper published by the standards organization ETSI [11]. Edge computing reduces latencies, mitigates communication jitter and enables mobility support [12]. Edge servers can execute local network management tasks in different scenarios. For example, they are used to realize distributed reputation management in vehicular networks [13] and promote decentralized electric vehicle discharging/charging management [14].

In this paper, we propose that in a region, a Local Authorities (LA) assigns edge servers to perform local data management tasks. With the help of LAs, different network entities are collaborative to accomplish network-wide data management, ranging from data transmission, storage, access, trading to mining. The advantages about the proposal for mobile healthcare systems are summarized as follows.

- *Real-time system response*: Healthcare data is transmitted to edge servers for prompt processing with lower response time. Users retrieve raw data, search data analysis results and participate in data trading on demand.
- *Balancing workload allocation*: Geo-distributed LAs cooperate with each other to process healthcare data synchronously when necessary. With workload balancing, scalable network management is supported to cope with the exponential data growth.
- *Reliable data protection*: In edge computing environment, data transmission, access and storage are localized. Healthcare data is easier to be protected well as data processing occurs in closer proximity to users [15].
- *Convenient data trading*: Pervasive connections provide data collectors and data owners convenience in carrying out data trading. The interaction among them also benefits from the coordination of edge servers.

To achieve the advantages, we present a secure and efficient data management system assisted by edge computing, named as EdgeCare, for mobile healthcare systems. We extend our work in [16], and focus on a new scheme adapting to edge computing paradigm but also study data trading. But designing a decentralized and collaborative data management system for healthcare applications is indeed a challenging task, especially when considering a large-scale mobile network environment and the practical issues of implementation.

We realize that critical issues should be addressed for guaranteeing security and efficiency of network implementation. First, architecture design is the fundamental step to ensure that related network entities are included and scheduled specifically for network-wide data management. Second, secure data uploading and customized access control are essential to enrich user satisfaction. To this end, protocol presentation with basic cryptographic techniques is required in the system. Last but not the least, for large-scale data trading, incentive mechanisms are necessary to reward data owners and stimulate them to participate in data trading.

Considering the specific challenging problems of data management in mobile healthcare systems, our work exhibits a practically viable solution from three systematic levels. More specifically, a hierarchical architecture is designed for wide application of EdgeCare. After that, we use electronic medical record (EMR) to design a use-centric EMR management scheme so that EMRs are uploaded and shared according to authorization granted by individual users. Furthermore, we consider a revenue maximization problem for an external data miner as a data collector that purchases access privileges of healthcare data from local users to execute a data mining task. The interaction between the data miner and users is formulated by Stackelberg game to approach feasible data trading environment.

We summarize the main contributions of this paper as follows.

- We propose the hierarchical architecture of EdgeCare to ensure the feasibility of secure and efficient data management in the decentralized mobile environment.
- We elaborately design the EMR management scheme and corresponding protocol workflows to handle healthcare data processing with security and privacy requirements.
- We conduct the Stackelberg game based optimization algorithm to approach the optimal incentive mechanism for the data collector and the users, which are involved in the fair decentralized data trading.

The rest of this paper is organized as follows. Section II presents the related work. In Section III, we propose the three-layer architecture for EdgeCare. We design the considerate EMR management scheme which supports secure data uploading and sharing in Section IV. As for the decentralized data trading scheme, we present the Stackelberg game model in Section V. Security analysis with numerical results are offered to demonstrate overall performance of EdgeCare in Section VI. Finally, Section VII concludes this paper.

II. RELATED WORK

A. CENTRALIZED DATA MANAGEMENT

Traditionally, centralized data management with cloud computing is first studied for mobile healthcare systems. Thilakanathan *et al.* [17] proposed a centralized cloud platform that allows doctors to monitor patients and share healthcare data with confidentiality. Similarly, the work in [18] considered utilizing a powerful cloud platform to collect, process and store healthcare data via unified standards. The platform is regarded as a fully trusted and central authority for all the network entities. Alternatively, Li *et al.* [19] focused on a semi-trusted cloud computing environment and tried to exploit attribute-based encryption techniques for achieving fine-grained data access control and avoiding privacy exposure to unauthorized parties. In particular, multiple authorities are employed to govern a disjoint subset of user role attributes.

As stated above, centralized data management cannot adapt to the increasing amount of the global healthcare data. In the schemes, overload states and intolerant processing delay are frequently resulted in when continuously processing a large volume of incoming healthcare data. We also find that due to the assigned core functionalities, the central authority poses potential security threats. The central authority is vulnerable to the attacks, including single point failure, DDoS attakcs and remote hijacking attacks.

B. BLOCKCHAIN-BASED DECENTRALIZED DATA MANAGEMENT

Recently, some studies focus on exploiting the blockchain technology to enable decentralized data management in mobile healthcare systems. Owing to transparency and shareability, blockchain is investigated to ensure data integrity and traceability for facilitating secure data access, sharing and storage in a decentralized manner. Reference [8] presented several healthcare use cases of blockchain to implement diverse healthcare services with security guarantee, e.g., public health surveillance and medication prescribing. The concept of blockchain in healthcare is growing interests. To cope with heterogeneous medical data, Kaur *et. al* [6] explored to combine blockchain with cloud computing for secure storage and sharing without involving third parties. The authors also emphasized the proposal still remains challenging as healthcare service providers, medical providers and organisations, public health agencies and governments need to work together and be encouraged with policy enforcement.

Furthermore, the work in [9] paid attention to developing a blockchain based access control for improving the migration of medical records to cloud-based platforms. More specifically, the permissioned blockchain technology was employed to permit anyone to access EMRs only after identity verification and membership authentication. As a consequence, data sharing was implemented with scalability, lower overheads and privacy protection. Similar work can be found in [7], which introduced a specific blockchain node and normal nodes with a veto power and general voting rights, respectively, to jointly determine whether uploading healthcare data is valid and acceptable. The goal was to ensure that any modifications to existing healthcare data can be validated and traced. Moreover, some prototype implementation about the blockchain-based systems were proposed subsequently, e.g., MedRec [10] and smart contract based clinical trials [20].

However, most of the blockchain-based schemes cannot support prompt data transfer and real-time data access for future mobile healthcare systems because of leveraging cloud storage. In addition, due to inherent complexity and potential risks about blockchain deployment, it is still a challenging task to smoothly implement practical systems even prototypes of the schemes in the real-world environment. Moreover, practical data trading is seldom investigated to build a healthcare data market in the researches.

Alternatively, we propose EdgeCare, where edge computing is integrated for decentralized data management and data trading to improve the overall system performance. Compared to blockchain-based healthcare systems, EdgeCare has predominant advantage in supporting large-scale data collection, ubiquitous data access and localized data processing. Besides, the deployment cost of EdgeCare could be moderate for the network operator.

III. HIERARCHICAL ARCHITECTURE

A. EDGE COMPUTING IN EDGECARE

To fulfill decentralized and collaborative data management, edge computing plays a critical role for the realization of EdgeCare. Due to proximal deployment of and seamless connection to edge servers, healthcare data is convenient to be gathered in local storage, promoting data collection and access with fleet response. Meanwhile, ubiquitous edge servers offer geo-distributed LAs with powerful managing capability to locally process healthcare data, ranging from preprocessing to analysis. If necessary, edge servers can also be scheduled as proxy servers of external data miners to negotiate with users as data owners in data trading, ultimately building a feasible and decentralized healthcare data market. We offer more details about the utilization of edge computing to facilitate users, LAs and data miners in EdgeCare as follows.

Owing to local storage, users retrieve the raw data within low latency. Users are encouraged to log in EdgeCare at any time and view the requests of data trading from data miners timely. Towards secure data trading, users can also assign edge servers to modify individual healthcare data and remove sensitive information in traded data. Then data trading is supported well to build a feasible healthcare data market, which realizes data utility and avoids privacy disclosure simultaneously.

By scheduling multiple edge servers, LAs are established and collaborative to manage network-wide healthcare data in a fully decentralized manner. Managing tasks are accomplished with on-demand resource provision. In EdgeCare, complex processing procedure caused by handling with healthcare data is shifted to edge servers, achieving fast response to users. Edge servers are proximal to collect, analyse and store valid healthcare data. Besides, supervised access control is easily enabled for permitting external entities, e.g., healthcare service providers and data miners, to access healthcare data with security considerations.

With authorized access to healthcare data, data miners run a variety of data mining algorithms in edge servers. The exploitation of edge servers is helpful to accelerate execution time and increase accuracy rate of data mining procedures. When negotiating with users in data trading, the decision-making capability of data miners can be enhanced by renting proxy edge servers to support the interaction with users.

B. THREE-LAYER ARCHITECTURE FOR EDGECARE

For elaborate architecture design, we introduce the motivation to design a mobile healthcare system with decentralized and collaborative data management, consisting of the following three folds. First, localized data storage should be realized for proximal data transfer and convenient data access. Second, direct and distributed data processing at the network edge is required to lower system response and balance managing workload. Third, necessary user interfaces can be designed well to improve user experience when they register and customize user-centric access control strategies.

To this end, we propose a hierarchical architecture for large-scale application of EdgeCare. As shown in Fig. 1, the architecture is compromised of three layers: user layer, edge layer and core layer. The layers work independently and they are also jointly connected to promote overall system performance. By using the architecture, a network operator is convenient to deploy functional network entities in different layers accordingly. They are uniformly scheduled for decentralized data management with high scalability, reliability and traceability. The essential healthcare data is orderly collected, separately processed with low latency, and authorized to access with permission. As a consequence, real-time system response, scalable data processing, reliable security protection and convenient data trading are supported, achieving network evolution for meeting the requirements of future mobile healthcare systems.

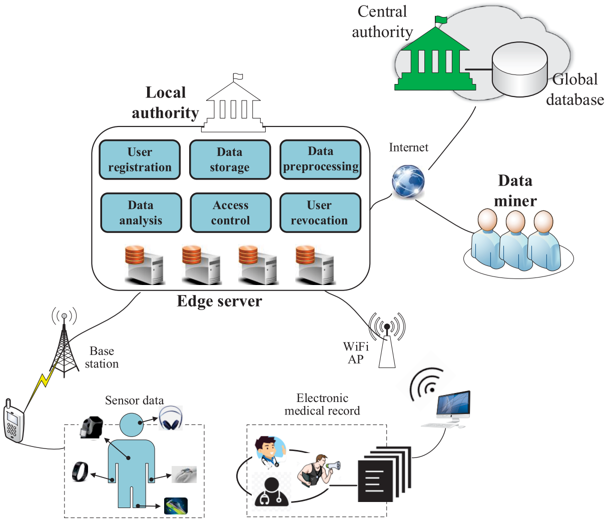


FIGURE 1. **Three-layer architecture for EdgeCare.**

In particular, the architecture can be applied to manage healthcare data for users belonging to diverse healthcare facilities, e.g., hospital and private clinic in several regions. Here, a region may range from a town to a city when the network operator determines the number and capability of all the LAs as a whole.

The presented architecture aims to orchestrate different network entities to cooperate with each other for collaborative data management. More specifically, in the bottom layer, different kinds of healthcare data are collected and may open to healthcare service providers for accessing and sharing. As for the middle layer, an LA employs multiple edge servers and communicates with local users and other LAs to support localized data processing. Meanwhile, data miners are allowed to access to healthcare data and perform data mining tasks after data trading with users. There exists a remote cloud in the top layer. A central authority resides in the remote cloud and holds complete identity information of all the network entities. The identity information with a whole mapping relationship table of the network entities is stored in a global database. It also supervises and coordinates geo-distributed LAs for cooperation when necessary. We provide more details about the hierarchical architecture with the network entities as follows.

1) USER LAYER

Nowadays, mobile healthcare systems support a variety of applications and numerous healthcare data is generated during service provision. On one hand, for personalized health guidance, a common user may simultaneously own serval healthcare service providers. Massive EMRs about the user are stored in different hospitals, clinics and medical institutions [21]. Besides, during the disease diagnosis and treatment, related personnel consisting of doctors, physicians and fitness instructors, set various electronic medical records.

The extensive adoption of wearable devices also promotes the consistent explosion of healthcare data. Diverse types of healthcare data is collected and uploaded when users utilize wearable devices to detect physical and mental health status. The advanced wearable devices generally include wrist devices, smart gloves and smart clothes. They measure heart rate, walking distance, body temperature, blood pressure and so on, and generate corresponding sensor data [22].

Ultimately, both the sensor data and EMRs are gathered together for further analysis in EdgeCare. They are packed and temporarily saved in portable devices (e.g., mobile phones and computers) via existing wired or wireless communication protocols, such as bluetooth, ZigBee and WiFi. When necessary, users transmit them to specified edge servers with the help of wiredly connected access points (e.g., base station and WiFi AP). Before being transmitted, raw healthcare data is processed for secure data transmission and reliable identity authentication. So cryptographic techniques such as encryption and signature are required in EdgeCare.

2) EDGE LAYER

With the establishment of LAs, crucial network functionalities of EdgeCare are concentrated to the network edge. Facing with large-scale healthcare data, the collaboration among several LAs is also proposed for a higher efficiency. On one hand, the whole network is divided into multiple regions and each LA has the managing scope to process uploaded healthcare data. In this way, total managing workloads are balanced. More healthcare data can be processed simultaneously. This greatly reduces overall delay, particularly when processing massive healthcare data. We further explain the rationale as follows.

- Proximal connection to adjacent LAs offers real benefit to reduce transmission latency for fast collecting and retrieving healthcare data.
- Multiple LAs cooperate with each other to tackle with network-wide data management, lessening queuing delay for local users in each region.
- Owing to the utilization of multiple LAs, EdgeCare becomes a typical multi-server queuing system. This advantage greatly lowers the potential congestion probability, and promotes the processing procedure with high reliability.

Moreover, when an LA is continuously busy for task execution, computing capability of local edge servers may not be sufficient. At this time, idle edge servers belonging to other LAs can be scheduled on demand to support the LA via existing techniques, e.g., remote procedure call. Under the circumstances, real-time communication and seamless cooperation among the LAs are exactly necessary.

As for a single LA, it can schedules edge servers to execute regular tasks about localized data management in a region. Supervised by the LA, healthcare data is shared to others including healthcare service providers and data miners. For data trading, data miners invite users to a healthcare data market. The data miners acquire access privileges granted by users and then are authorized to apply mining algorithms to healthcare data for extracting useful information. To sum up, the LA supports regular operations of users, healthcare service providers and data miners as follows:

- *User Register*: Before joining EdgeCare, users should register with valid identities, which are used for identifying the owner of healthcare data. After legal registration, an LA acts as trusted authority to generate and distribute public/private key pairs, digital signature and certificates for local users. The registration information is uploaded to the central authority in the remote cloud.
- *Data Storage*: As as a data repository in a region, the LA stores incoming raw healthcare data in local storage according to specified compressed and encrypted format. Besides, history information about data transmission, data processing, data access and data storage is recorded orderly. By searching the information, the traceability of happening events in EdgeCare is ensured. The system is convenient to investigate once any cases caused by security attacks and privacy leakage appear.
- *Data Preprocessing*: Raw healthcare data is collected and should be preprocessed first. In EdgeCare, appropriate data preprocessing is necessary due to various data sources, formats and structures of receiving healthcare data from different healthcare applications. For data miners, robust and complex approaches about data cleaning, integration, reduction and normalization can be implemented well by assigning edge servers. Moreover, when data miners are granted to access to healthcare data, the LA also helps modify data to hide sensitive information for users, instead of directly releasing data to data miners.
- *Data Analysis*: In edge computing environment, healthcare service providers and data miners are of great capability for processing massive healthcare data dynamically. They can combine real-time analysis and offline analysis according to different scenarios. In intensive care, collecting healthcare data should be processed quickly. Analysis results are forwarded to related personnel for low-latency response to emergency situations. As for those services with ever-lasting monitoring procedure, e.g., health planning and medical recommendation, offline analysis may be adequate for service provision in EdgeCare.
- *Access Control*: To improve access efficiency, EdgeCare provides a well-designed and user-centric interface for local users to achieve full control of their healthcare data. In EdgeCare, users can independently assign/revoke access privileges for/from others. After data trading,

as the owner of healthcare data, users customize and update flexible access policies for all the related personnel, e.g., data miners, based on the relationships with them. According to given access policies, the LA checks whether there exist unauthorized access to healthcare data.

- *User Revocation*: Revocation operation is activated in special situations. For example, one user has malicious behavior in the system and is punished by the LA to remove its account or one user would like to be managed by other LAs. Such an on-demand user revocation operation is to support necessary change of management ownership. In particular, the LA feedbacks a revocation notification to both the user and the central authority.

3) CORE LAYER

The central authority acts as the top security manager and is responsible for ensuring network-wide security protection. In practice, the central authority is a public and authentic organization fully trusted by all the network entities. With the highest priority of data management, the central authority records and accesses identity information about all the network entities in EdgeCare. The central authority also records the corresponding mapping relationships among them in mobile healthcare systems. The global database is built to maintain and update a large number of recording information.

In addition, via the Internet, the central authority acts as a network orchestrator to monitor whether each LA works normally and coordinate them for required cooperation when necessary. The functionality of network supervision and coordination are executed in a pre-active way. This means that all the responses and actions of the LAs are triggered by a preprogrammed global module in the central authority.

The involvement of the core layer significantly improves the working efficiency of the system, especially in prompt response to fault diagnosis and network attacks. State supervision for the LAs is supported to ensure each LA works safely. If an LA becomes faulted, the LA could be rapidly identified and replaced in time. Impersonation attacks caused by compromised LAs would be timely removed to overcome security threats. After identity authentication and real-time monitoring, behavior of all the LAs is of transparent trustworthiness. In short, with the presence of the core layer, the geo-distributed LAs are able to manage healthcare data in different regions in a highly efficient and intelligent way.

To perform the above operations, the central authority should be equipped with tamper-resistant and powerful hardware. So it is generally deployed in the remote cloud, and thus utilizes sufficient cloud resources to accomplish the tasks.

IV. SECURE DATA UPLOADING AND SHARING

Towards pervasive healthcare services, a key problem of EdgeCare is how to enable secure data uploading and sharing. In this paper, we consider basic protocol design for EdgeCare when users upload their helathcare data and share the data with others. For simplicity, we focus on a daily scenario that a patient authorizes several doctors to access his EMRs. EMR is a special healthcare data for patients and it is vital to guarantee EMR uploading and sharing with security and privacy protection in EdgeCare. Thus, we propose a considerate EMR management scheme wherein EMRs are uploaded and shared in a strictly secure way. To summarize, the scheme is compromised of four stages, as shown in Fig. 2. The notations used in the scheme are listed in Table 1.

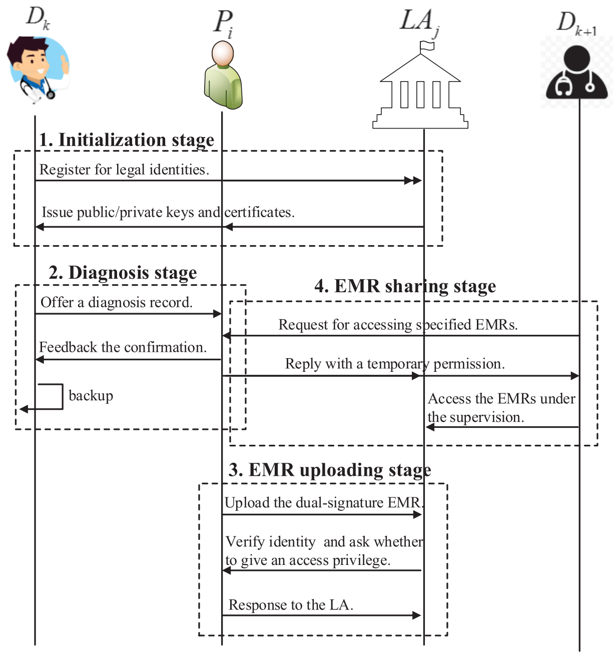

D_k
P_i
LA_j
D_{k+1}
1. Initialization stage
Register for legal identities.
Issue public/private keys and certificates.
2. Diagnosis stage
Offer a diagnosis record.
Feedback the confirmation.
backup
4. EMR sharing stage
Request for accessing specified EMRs.
Reply with a temporary permission.
Access the EMRs under the supervision.
3. EMR uploading stage
Upload the dual-signature EMR.
Verify identity and ask whether to give an access privilege.
Response to the LA.

FIGURE 2. EMR management scheme for EdgeCare with security guarantee.

TABLE 1. Notations and descriptions.

Notation	Description
P_i, LA_j and D_k	The i_{th} patient, j_{th} local authority and k_{th} doctor.
PK_e, SK_e and $Cert_e$	The owned public key, private key and certificate of entity e.
$i \rightarrow j : m$	Entity i transmits message m to entity j.
$E_{PK_e}(m)$	Encrypt message m with public key of entry e.
$Sign(SK_e, m)$	Sign message m with private key of entry x.
$\Vert Cert_e$	The certificate of entity e is attached.
SN	Treatment serial number created by a doctor.
$hash()$	A function used to obtain the hash value.

A. INITIALIZATION STAGE

A patient P_i and two doctor D_k, D_{k+1} first apply to the system for valid registration. After registration, they are issued with a public/private key pair and legal certificate, respectively. The identity information is also uploaded by an LA to the central authority for information record.

B. DIAGNOSIS STAGE

D_k diagnoses P_i and creates the corresponding EMR to P_i. The diagnosis-related data with the doctor's identity, hospital name (Hname), department name (Dname), and diagnosis time is saved into the EMR. The patient confirms the generated EMR and finally, the doctor records the EMR for backuping. More specifically, the steps related to the data processing procedure are as follows.

- *Step 1*: After diagnosis, D_k generates the diagnosis record (indicated by dr) with specified content and treatment serial number (SN). For authentication, D_k signs the diagnosis record and shows the signed result with the certificate to P_i. The result is encrypted with the public key PK_{P_i} and attach with the current time slot $Timestamp$.

$$dr = \{content||SN\},$$

where $SN = hash(Hname||Dname||D_k||P_i)$.

$$\begin{aligned} &D_k \rightarrow P_i : \\ &result = E_{PK_{P_i}}(Sign(SK_{D_k}, dr)||Cert_{D_k}||Timestamp) \end{aligned}$$

- *Step 2*: P_i decrypts the result, checks the diagnosis record and verifies the signature from D_k. P_i confirms the result to signs it and replies to the doctor.

$$\begin{aligned} &P_i \rightarrow D_k : \\ &reply = E_{PK_{D_k}}(Sign(SK_{P_i}, Sign(SK_{D_k}, dr))||Cert_{P_i} \\ &\qquad ||Timestamp). \end{aligned}$$

Finally, the EMR with dual signature is sent by the patient to the LA for data update.

- *Step 3*: D_k receives the reply, and then uploads a mapping list to private servers of the hospital for backuping. This mapping list consists of the identity of doctor and patient, department name and the treatment serial number:

$Dname$	D_k	P_i	SN

.

C. EMR UPLOADING STAGE

We assume that the patient belongs to LA_j. The EMR uploading stage is introduced as follows.

- *Step 1*: P_i encrypts the EMR with dual signature by using public key of LA j, and sends it to the LA.

$$\begin{aligned} &P_i \rightarrow LA_j : \\ &EMR = E_{PK_{LA_j}}(Sign(SK_{P_i}, Sign(SK_{D_k}, dr))||Cert_{P_i} \\ &\qquad ||Timestamp). \end{aligned}$$

- *Step 2*: LA_j realizes the arrival of a new EMR and decrypts it to verify legal identity of the patient first. Then LA_j realizes the EMR has been signed by the user, and stores the EMR locally but also updates information record of the healthcare data. Meanwhile, LA_j finds the attached signature from D_k. LA_j searches whether the doctor has been assigned any access privileges by the patient. When there exists no, the LA sends a query to the patient and ask whether the doctor should own a certain access privilege from him.
- *Step 3*: Based on individual purpose, the patient can refuse to assign or directly set a given access privilege for the doctor by logging into EdgeCare.

D. EMR SHARING STAGE

EMRs are critically important information for diagnosis and treatment in moible healthcare systems. The information is also private for individual. And the private information are generally required to share among different medical personnel for improving the quality of healthcare services [23]. In this paper, we aim to facilitate EMR sharing in EdgeCare with necessary security considerations.

When the patient is treated by a new doctor D_{k+1}, the doctor may review a set of EMRs regarding to P_i for further diagnosis. But the patient has not ever assigned any access privileges for the doctor. The doctor should be permitted by the patient to access to the healthcare data temporarily. At this time, EMR sharing procedure is shown as follows.

- *Step 1*: D_{k+1} sends a request to the patient for accessing to a set of EMRs $\{EMR\}^{P_i}_{D_{k+1}}$ within the valid time. The accessing request from D_{k+1} is denoted as $AQ(\{EMR\}^{P_i}_{D_{k+1}}, validtime)$. We express the accessing request transmitted from the doctor to the patient as follows.

$$\begin{aligned} &D_{k+1} \rightarrow P_i : \\ &request = E_{PK_{P_i}}(Sign(SK_{D_{k+1}}, AQ(\{EMR\}^{P_i}_{D_{k+1}}, \\ &\qquad validtime))||Cert_{D_{k+1}}||Timestamp). \end{aligned}$$

- *Step 2*: P_i receives the request and also verifies the signature from D_{k+1}. The patient realizes the legal identity of the doctor and independently chooses whether to offer the permission. When the patient gives out a positive permission, two replies including $perm1$ and $perm2$ are fed back to D_{k+1} and LA_j, respectively.

$$\begin{aligned} &P_i \rightarrow D_{k+1} : \\ &perm1 = E_{PK_{D_{k+1}}}(Sign(SK_{P_i}, Sign(SK_{D_{k+1}}, AQ(\\ &\qquad \{EMR\}_{D_{k+1}}, validtime))||Cert_{P_i}||Timestamp) \\ &P_i \rightarrow LA_j : \\ &perm2 = E_{PK_{LA_j}}(Sign(SK_{P_i}, Sign(SK_{D_{k+1}}, AQ(\\ &\qquad \{EMR\}^{P_i}_{D_{k+1}}, validtime))||Cert_{P_i}||Timestamp). \end{aligned}$$

- *Step 3*: When the permission $perm1$ is received by the doctor, D_{k+1} requests for accessing to specified EMRs ($\{EMR\}^{P_i}_{D_{k+1}}$) stored in LA_j. The external request er is expressed as follows.

$$\begin{aligned} &D_{k+1} \rightarrow LA_j : \\ &er = E_{PK_{LA_j}}(Sign(SK_{D_{k+1}}, AQ(\{EMR\}^{P_i}_{D_{k+1}}, \\ &\qquad validtime))||Cert_{D_{k+1}}||Timestamp). \end{aligned}$$

LA_j realizes that the doctor has been granted by P_i with the verification of dual signature in $perm2$. Thus, the LA admits that D_{k+1} has temporary privilege to read the EMRs of P_i. In the meantime, the LA also supervises whether the temporary access follows the principle of presetting privilege within the given time span. Overall, the EMRs are shared to the legitimate doctor with accountability in EdgeCare.

V. DECENTRALIZED DATA TRADING

A. BASIC PRINCIPLES IN DECENTRALIZED DATA TRADING

The primary concern for users in data trading is how to avoid privacy violation and get enough compensations when releasing individual healthcare data to data miners. On one hand, users should take action to make their individual data only accessible to data miners with granted access privileges. Besides, sensitive information need to be protected well for privacy preservation. For users, sensitive information attached to raw healthcare data may consist of users' name, ID card numbers, cell phone number and so on [24]. Nowadays, for data anonymity, desensitization is widely used to remove or modify identifiers and quasi-identifiers [25]. In this way, sensitive information is excluded in data publishing for achieving privacy-preserving data mining.

As for a data miner, it would like to collect interesting healthcare data from users for performing a data mining task. To accomplish a data collection task across the whole network, the data miner may rent distributed edge server as the proxy server to negotiate with users for carrying out data trading in each region. This exactly promotes decentralized data trading. At the same time, the data miner needs to reward local users as data owners with incentive. The incentive may be regarded as monetary compensations for potential risks in privacy disclosure and economic income to stimulate users in data trading [26].

According to the above principles, decentralized data trading between users and data miners is facilitated by utilizing edge servers in EdgeCare. First, a trusted edge server in an LA acts as a broker of a data miner and realizes the request about data collection. The proxy edge server helps seek appropriate data owners associated with interesting data. Second, local users mandate edge servers to prevent unauthorized access and perform data modification operations for desensitization before publishing healthcare data to the data miner. In the meantime, edge servers notify data owners with the related information about the request of data trading by proximally communicating with them. Edge servers help transmit reward policy to the data owners. Finally, interesting healthcare data is gathered by the proxy edge server. The data miner can directly assign the proxy edge servers for data mining and take the data mining results to individual servers [27]. In short, edge servers are crucial for coordinating decentralized data trading among both sides of users and the data miner in EdgeCare.

B. STACKELBERG GAME APPROACH

An incentive mechanism is necessary when encouraging users to participate in data trading. In the incentive mechanism, reward policy is set for stimulating data owners managed by an LA to offer access privileges about their healthcare data. Facing with various reward policies, users response with different participation levels in data trading according to their acquired utilities.

Next, we formulate a utility function of one user in data trading. The utility function is related to a trade-off between privacy gains and economic benefits resulted by participating in data trading [28]. We consider that a user i owns several data records interested by a data miner j. The user becomes an appropriate data provider for the data miner in the region. The proxy edge server represents the data miner to send a request of data trading to the user. User i needs to determine the participation level in terms of the amount of data records granted to data miner j, which is denoted as x_i^j. x_i^j is a vital decision variable to influence the final utility of user i.

When the user contributes more data records to the data miner, user i clearly obtains more rewards, $x_i^j r_i^j$. r_i^j is the rewards given by data miner j for accessing to a data record of user i. However, with the increasing amount of traded data, the risk of privacy disclosure is increased potentially. The increasing risk means that a compromised data miner may further infer, estimate and recover sensitive information based on more released data records [29]. This leads to privacy losses for the user. In turn, user i obtains privacy gains by reducing the amount of data records accessed by the data miner. To summarize, the utility function of user i consists of two aspects: subjective privacy gains and actual economic benefits, which is expressed by

$$U_i^j = w_i \log(X_i - x_i^j + 1) + r_i^j x_i^j,\ \ r_i^j \geqslant \rho_i d_j \tag{1}$$

Here, w_i is the individual willingness for strict data protection rather than data trading [30]. The value of w_i is related with the specified attitude for the kind of data records. X_i is the total amount of data records belonging to the user interested by data miner j. Hence, $w_i \log(X_i^j - x_i^j + 1)$ is the so-called privacy gains consciously believed by user i when permitting data miner j to access to a part of data records. d_j is the current market price of each data record interested by the data miner. In a big healthcare data market, the value of d_j is mainly related with the kind (e.g., dimensionality and attribute) of accessing healthcare data, and may be published to all the users [31]. ρ_i represents individual preference of user i for data trading after evaluating data value of the owned healthcare data. Clearly, ρ_i is subjectively increased when the user regards the owned healthcare data with higher value. There exists a user requirement, $r_i^j \geqslant \rho_i d_j$.

For the data miner, its utility function is directly bound up with the final revenue R_j in data trading. Data miner j aims to access to enough data records for maximizing the economic benefits. For the data miner, the income earned by mining these data records is equal to $(1 + t_j)d_j \sum_{i\in\mathcal{I}} x_i^j$, where t_j is the capability of data miner j in transforming data values to economic assets after performing the data mining task. As for the total payments, they are represented by $\sum_{i\in\mathcal{I}} x_i^j r_i^j$. Here, $\mathcal{I}$ is the set of users that own healthcare data interested by the data miner. Then $R_j = (1 + t_j)d_j \sum_{i\in\mathcal{I}} x_i^j - \sum_{i\in\mathcal{I}} x_i^j r_i^j$.

Besides, there exist several constraints when data miner j tries to get authorized access to interesting healthcare data of users belonging to the set $\mathcal{I}$. For each member, the amount of data records granted to data miner j should be within the limitation of an upper value, $y^j_{\max}$. In general, the constraint is for ensuring heterogeneous data sources in data mining [32]. To acquire sufficient data records for getting achievable data utility in data mining, the total amount of data records in the region offered to data miner j is also required to be larger than a threshold value, $Y^j_{\min}$. Moreover, the data miner needs to provide an attractive reward policy for users belonging to the set $\mathcal{I}$ compared with the market price, d_j.

Based on the considerations, we try to make the mathematic model exactly adapt to the feasible data trading environment. As a consequence, this gives rise to a revenue maximization problem with serval constraints for data miner j when purchasing access privileges from local users. The problem is presented as follows.

$$\begin{aligned} &\max_{r^j_i, x^j_i} R_j \\ &\text{s.t. } 0 \le x^j_i \le \min(X_i, y^j_{\max}) \\ &\qquad \sum_{i \in \mathcal{I}} x^j_i \ge Y^j_{\min} \\ &\qquad r^j_i \ge \rho_i d_j \end{aligned} \tag{2}$$

where $\min(n_1, n_2)$ is a function to obtain the minimum value among two given numbers, n_1 and n_2.

To solve the problem, we formulate the interaction between a data miner (data collector) and multiple users (data owners) in decentralized data trading by using Stackelberg game approach. Data miner j offers monetary incentive indicated by r^j_i to each user when requesting for accessing to the healthcare data. In turn, user i replies to the request from the data miner with its participation level based on the announced rewards. Namely, the user determines the amount of data records granted to data miner j, x^j_i. For user i, the determination is optimized to maximize the utility, which is related to a privacy-incentive trade-off function in Eqn. (1). All the responses from users are collected to data miner j for assisting in decision making. Then the data miner further adjusts reward policy for various users of different responses in order to maximize the revenue with meeting necessary constraints, as shown by the problem in (2).

This means that the data miner is naturally fit for acting as a leader to determine final reward policy while all the users become followers responding to the data miner with respect to given rewards. Thus, the interaction between the data miner and all the users can be formulated as a typical two-stage leader-follower game in the incentive mechanism. According to the descriptions in [33], Stackelberg game model is a convenient analytical model to study the above scenario. We define the Stackelberg game between data miner j and all the users by the following strategic formčž

$$\Gamma = \{(j \cup \{i\}_{i\in\mathcal{I}}), R_j, \{U^j_i\}_{i\in\mathcal{I}}, \{r^j_i\}_{i\in\mathcal{I}}, \{x^j_i\}_{i\in\mathcal{I}}\}.$$

In particular, $(j \cup \{i\}_{i\in\mathcal{I}})$ is the player set of the game Γ. R_j and $\{U^j_i\}_{i\in\mathcal{I}}$ are utility sets of the leader and followers, respectively. For decision making, $\{r^j_i\}_{i\in\mathcal{I}}$ and $\{x^j_i\}_{i\in\mathcal{I}}$ are their strategy sets.

In EdgeCare, the Stackelberg game with a single leader and multiple followers is performed well to implement decentralized data trading. The data miner carries out data trading in different regions to finish the whole data collection task. In each region, data miner j has the privilege to act first while the followers react according to the action of the leader. For revenue maximization, the proxy edge server in an LA becomes a broker of data miner j, which tells the proxy edge server about the data collection requirements, including t_j, $y^j_{\max}$ and $Y^j_{\min}$. The proxy edge server formulates and solves problem in (2) to determine the optimal reward parameter p^{j*}_i based on prior knowledge about the impacts of the decisions on behavior of followers.

Here, we consider that due to attractive incentive, users are willing to upload private information (w_i, X_i and ρ_i) to the edge server. The proxy edge server represents data miner j to declare the rewards p^{j*}_i for encouraging users. To maximize the individual utility, the best participation level in terms of x^{j*}_i is given by user i when replying to data miner j in data trading. The data miner handles the proxy edge server to access to the corresponding healthcare data and gather them for performing the data mining task. More details about the practical scenario regarding to the data miner, proxy edge server and users are shown in Fig. 3.

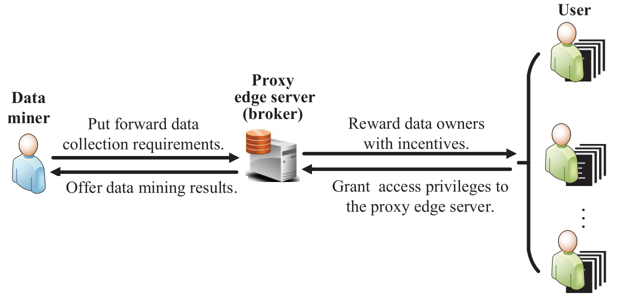

User
Data miner
Proxy edge server (broker)
Put forward data collection requirements.
Offer data mining results.
Reward data owners with incentives.
Grant access privileges to the proxy edge server.

FIGURE 3. A broker model for decentralized data trading between a data miner and local users in EdgeCare.

C. STACKELBERG EQUILIBRIUM ANALYSIS

The objective of the proposed Stackelberg game Γ between data miner j and user i is to find the unique Stackelberg equilibrium where both of them have no motivations to unilaterally change their decisions. Next, we exploit the Stackelberg game theory approach to analyse the best responses of the leader and followers in data trading, respectively.

We study the best responses of the followers first. U^j_i can be converted into an optimal utility function in terms of x^j_i. We take the first and second derivatives of U^j_i with respect to x^j_i, which are shown by

$$\begin{aligned} \frac{\partial U^j_i}{\partial x^j_i} &= -\frac{w_i}{X_i - x^j_i + 1} + r^j_i \\ \frac{\partial^2 U^j_i}{\partial x^{j2}_i} &= -\frac{w_i}{(X_i - x^j_i + 1)^2} < 0. \end{aligned} \tag{3}$$

We easily find that $\partial^2 U^j_i / \partial x^{j2}_i < 0$. The utility function is concave. This means that there exists the maximal value of U^j_i

and the optimal solution of x_i^j, denoted as x_i^{j*}, can be achieved. We obtain x_i^{j*} by using $\partial U_i^j/\partial x_i^j = 0$, then have

$$x_i^{j*} = X_i - \frac{w_i}{r_i^j} + 1. \tag{4}$$

For user i, x_i^{j*} is called the best response under given rewards r_i^j, which maximizes individual utility on the condition of the incentive from data miner j. Clearly, when combining feasible constraints, the best response of the user is further stated in practice as follows

$$x_i^{j*} = \begin{cases} 0, \; r_i^j < \rho_i d_j \\ X_i - \frac{w_i}{r_i^j} + 1, \quad \rho_i d_j \leqslant r_i^j \leqslant w_i \\ X_i, \; r_i^j \geq w_i \end{cases} \tag{5}$$

As a leader in the game, the data miner can know x_i^{j*} from each user after holding the prior knowledge about the impacts of the decisions on behavior of the followers. We substitute x_i^{j*} into the revenue maximization problem of data miner j in (2), which is reformulated as

$$\begin{aligned} &\max_{r_i^j} \sum_{i\in\mathcal{I}} [(d_j + t_j d_j - r_i^j)(X_i + 1) - \tfrac{(d_j+t_j d_j)w_i}{r_i^j} + w_i] \\ &\text{s.t. } \rho_i d_j \leqslant r_i^j \leqslant r_i^{\max} \\ &\quad \sum_{i\in\mathcal{I}} \tfrac{w_i}{r_i^j} \leqslant \sum_{i\in\mathcal{I}} X_i + |\mathcal{I}| - Y_{\min}^j \end{aligned} \tag{6}$$

Here, $r_i^{\max} = \frac{w_i}{X_i+1-\min(X_i, y_{\max}^j)}$ and $|\mathcal{I}|$ indicates the number of members in the set $\mathcal{I}$. We simplify the above problem with an intermediate variable $r_i^{j'} = \frac{w_i}{r_i^j}$ as follows

$$\begin{aligned} &\min_{r_i^{j'}} \sum_{i\in\mathcal{I}} [\tfrac{w_i(X_i+1)}{r_i^{j'}} + (d_j + t_j d_j) r_i^{j'}] \\ &\text{s.t. } a_i \leq r_i^{j'} \leq b_i \\ &\quad \sum_{i\in\mathcal{I}} r_i^{j'} \leqslant c \end{aligned} \tag{7}$$

where $a_i = \frac{w_i}{r_i^{\max}}$, $b_i = \frac{w_i}{\rho_i d_j}$ and $c = \sum_{i\in\mathcal{I}} X_i + |\mathcal{I}| - Y_{\min}^j$.

By proposing the Lagrange multipliers α_i, β_i and γ for the constraints accordingly, the Lagrange function of the simplified problem L is expressed by

$$\begin{aligned} L = \sum_{i\in\mathcal{I}} [\frac{w_i(X_i+1)}{r_i^{j'}} + (d_j + t_j d_j) r_i^{j'}] - \sum_{i\in\mathcal{I}} \alpha_i (r_i^{j'} - a_i) \\ + \sum_{i\in\mathcal{I}} \beta_i (r_i^{j'} - b_i) + \gamma (\sum_{i\in\mathcal{I}} r_i^{j'} - c). \end{aligned} \tag{8}$$

We take the first and second derivatives of L with respect to $r_i^{j'}$, and find

$$\begin{aligned} \frac{\partial L}{\partial r_i^{j'}} &= -\frac{w_i(X_i+1)}{r_i^{j'2}} + d_j + t_j d_j - \alpha_i + \beta_i + \gamma \\ \frac{\partial^2 L}{\partial r_i^{j'2}} &= \frac{2 w_i (X_i+1)}{r_i^{j'3}} > 0 \end{aligned} \tag{9}$$

Clearly, the optimization problem in (7) is convex with linear constraints. An optimal solution of $r_i^{j'}$ is solved. Furthermore, there exists a unique r_i^{j*} to be acquired as the best response of data miner j after learning all the responses of x_i^{j*}.

In this paper, we use primal-dual Lagrangian method to solve the problem in (7). According to the first-order derivative optimality condition, we have

$$r_i^{j'*} = \sqrt{\frac{w_i(X_i+1)}{d_j + t_j d_j - \alpha_i + \beta_i + \gamma}}. \tag{10}$$

Besides, we also get the following complementary slackness condition based on the KKT condition

$$\begin{aligned} \alpha_i (r_i^{j'} - a_i) &= 0 \\ \beta_i (r_i^{j'} - b_i) &= 0 \\ \gamma (\sum_{i\in\mathcal{I}} r_i^{j'} - c) &= 0 \\ \alpha_i, \beta_i, \gamma &\geqslant 0 \end{aligned} \tag{11}$$

We consider possible cases as follows.

Case 1: $\alpha_i = 0$, $\beta_i = 0$ and $\gamma = 0$. According to Eqn. (10), $r_i^{j'*} = \sqrt{\frac{w_i(X_i+1)}{d_j+t_j d_j}}$. We also ensure that $\sum_{i\in\mathcal{I}} r_i^{j'} \leqslant c$ and $a_i \leqslant r_i^{j'*} \leqslant b_i$. Hence, $p_i^{j'*}$ is updated as follows:

$$r_i^{j'*} = \begin{cases} a_i, \sqrt{\frac{w_i(X_i+1)}{d_j + t_j d_j}} < a_i \\ \sqrt{\frac{w_i(X_i+1)}{d_j + t_j d_j}}, \quad a_i \leqslant \sqrt{\frac{w_i(X_i+1)}{d_j + t_j d_j}} \leqslant b_i \\ b_i, \sqrt{\frac{w_i(X_i+1)}{d_j + t_j d_j}} > b_i \end{cases} \tag{12}$$

Case 2: $\alpha_i = 0$, $\beta_i = 0$ and $\gamma \neq 0$. There exist $r_i^{j'*} = \sqrt{\frac{w_i(X_i+1)}{d_j+t_j d_j+\gamma}}$ and $\sum_{i\in\mathcal{I}} r_i^{j'} = c$. Thus, we easily obtain

$$\gamma = \frac{(\sum_{i\in\mathcal{I}} \sqrt{w_i(X_i+1)})^2}{c^2} - (1 + t_j) d_j. \tag{13}$$

We further calculate $r_i^{j'*}$ as follows

$$r_i^{j'*} = \frac{c\sqrt{w_i(X_i+1)}}{\sum_{i\in\mathcal{I}} \sqrt{w_i(X_i+1)}}. \tag{14}$$

Similar, we should guarantee $a_i \leqslant r_i^{j'*} \leqslant b_i$. We update $r_i^{j'*}$ based on the following algorithm with iterative steps:

Step 1: Initialization. The set $\mathcal{I}$, boundary values a_i, b_i and c, and an empty set $\mathcal{A}$ are initialized.

Step 2: Operation. For each member belonging to $\mathcal{I}$, calculate $r_i^{j'*}$ according to Eqn. (14). If the value of $r_i^{j'*}$ is beyond the boundary limitation, $r_i^{j'*}$ is replaced with the nearest boundary value. The member is also recorded into $\mathcal{A}$.

Step 3: Termination. If $\mathcal{A}$ is still empty, the algorithm is terminated. Otherwise, $\mathcal{I}$, c and $\mathcal{A}$ are updated: $\mathcal{I} = \mathcal{I} \backslash \mathcal{A}$, $c = c - \sum_{i \in \mathcal{A}} r_i^{j\prime *}$ and $\mathcal{A} = \emptyset$, then jump to Step 2. Here, $\mathcal{I} \backslash \mathcal{A}$ means an operation to delete the elements owned by $\mathcal{A}$ from $\mathcal{I}$ and $\emptyset$ represents an empty set.

Case 3: $\alpha_i \neq 0$, $\beta_i = 0$ and $\gamma = 0$. There exist $r_i^{j\prime *} = a_i$ and $\sum_{i \in \mathcal{I}} r_i^{j\prime} \leqslant c$. This means that $\sum_{i \in \mathcal{I}} a_i \leq c$ in this case.

Case 4: $\alpha_i \neq 0$, $\beta_i = 0$ and $\gamma \neq 0$. There exist $r_i^{j\prime *} = a_i$ and $\sum_{i \in \mathcal{I}} r_i^{j\prime} = c$. So we infer that the requirement $\sum_{i \in \mathcal{I}} a_i = c$ should be satisfied at this time.

Case 5: $\alpha_i = 0$, $\beta_i \neq 0$ and $\gamma = 0$. There exist $r_i^{j\prime *} = b_i$ and $\sum_{i \in \mathcal{I}} b_i \leqslant c$.

Case 6: $\alpha_i = 0$, $\beta_i \neq 0$ and $\gamma \neq 0$. There exist $r_i^{j\prime *} = b_i$ and $\sum_{i \in \mathcal{I}} b_i = c$.

By using the primal-dual Lagrangian method, a proxy edge server can reach the Stackelberg equilibrium (x_i^{j*}, r_i^{j*}) in this paper. As stated above, the proxy edge server holds the knowledge of the related information of both data miner j and all the users in data trading. The proxy edge server acts as a broker to transform the original problem in (2) to the problem in (6), and find the optimal solution of $r_i^{j\prime}$ by solving the simplified problem in (7). Then the proxy edge server further calculates r_i^{j*}. After that, x_i^{j*} is presented by each user as predicted when facing with r_i^{j*}.

Theorem 1: A unique Stackelberg Equilibrium exists between the data miner and all the users in the proposed Stackelberg game model.

Proof: In data trading, according to the given reward parameter r_i^j, each user acts as a follower and always has its own best response x_i^{j*}. x_j^{i*} is unique due to the concave character of the utility function, namely, $\partial^2 U_i^j / \partial x_i^{j2} < 0$, as shown in Eqn. (3). By having insight into all the best responses x_i^{j*}, $\forall i \in \mathcal{I}$, the revenue maximization problem of the leader is reformulated accordingly. Then r_i^{j*} is solved via the proposed primal-dual Lagrangian method.

Meanwhile, we demonstrate that the data miner has a unique optimal strategy under given the best strategies of all the users. Ultimately, in the game model, both the leader and followers are fully satisfied because their decisions make that their utilities have been maximized simultaneously. Moreover, when all the players, including each user and the data miner, have their optimized payoff and cost, respectively, considering the strategies chosen by other players in the game. They have no incentives to change the decisions and take other actions. Thus, (r_i^{j*}, x_i^{j*}) is obtained to guarantee that the unique Stackelberg equilibrium is reached finally. ■

VI. PERFORMANCE EVALUATION

A. SECURITY ANALYSIS FOR PROTOCOL DESIGN

We offer necessary security analysis to demonstrate that EdgeCare with basic protocols has great advantages in security and privacy protection for users in decentralized and collaborative data management. In this paper, an EMR management scheme is to take EMR as an example for showing that how healthcare data is processed with elaborative security considerations. According to the descriptions in [34], some important security and privacy requirements are exactly satisfied in our proposed EdgeCare. We summarize the significant advantages of EdgeCare into the following aspects.

1) CONFIDENTIALITY

In EdgeCare, confidentiality of communications is protected by exploiting the standard cryptographic primitives. We utilize asymmetric/symmetric key-based encryption and digital signatures in our schemes. Without the symmetric keys and private keys of any entities, any potential adversaries cannot open the encrypted packets even though they may realize the existence of packets and steal them by eavesdropping on wireless communications and illegal packet capture. During the communications, we use timestamp in all the packets to effectively prevent replay attacks.

2) INTEGRITY AND AUTHENTICATION

In the EMR management scheme, after treatment, a doctor must sign the diagnosis record. The diagnosis record with the digital signature is sent to a patient. By verifying the digital signature, the patient confirms the diagnosis record and further signs it. Thus, the diagnosis record with dual signature is finally generated to reach a consensus. Then the mutually agreed EMR is transferred for data update in the LA. Here, without the private key of the signer, any entity cannot counterfeit the digital signature of other entities. Since the digital signature is only generated by a specific signer, any information with a digital signature can be authenticated and verified whether the signer is the sender or not. If one EMR is modified by unauthorized parties or random errors during the transmission, the receivers can also discover it in the process of verification, which guarantees the integrity and authentication.

3) TRACEABILITY

Due to the proposed dual signature, non-repudiation of designed communication protocols is ensured. In case of a round of packet transmission and receipt, neither the sender nor the receiver can deny having taken part in the communication. This means that the communication protocols are able to avoid one of the implied entities (i.e., sender and receiver) cheating and being cheated. At the same time, traceable data access is supported as LAs record and monitor who access to healthcare data of local users.

4) USER-CENTRIC ACCESS CONTROL

In the EMR management scheme, when a doctor would like to access the healthcare data of a patient, the doctor should ask the patient for permission. Otherwise, the LA will prevent the access caused by the doctor. Similarly, when the doctor hopes to acquire a regular access privilege from the patient, the doctor should also directly inquire the patient whether it can be granted. In short, for the patient, all the data access in the LA regarding to his healthcare data should be firstly authorized by him. Moreover, the patient can independently permit temporary access and assign/revoke any access privileges for/from others in EdgeCare.

B. NUMERICAL RESULTS ABOUT DECENTRALIZED DATA TRADING

We evaluate the performance of the proposed Stackelberg game based scheme for decentralized data trading in EdgeCare by extensive simulations. A data miner tries to access to healthcare data belonging to several users for performing a data mining task. For simplicity, we consider that there exist 100 users owning healthcare data interested by the data miner in a region. The amount of healthcare data belonging to each user is distributed uniformly over [1, 5] kilo records. For the users, the individual willingness for data protection rather than data trading w ranges from 3 to 8. For feasible data mining, data collection requirements are put forward, including t, $y_{\max}$ and $Y_{\min}$. They are 1, 4 and 200 kilo records. Individual preference ρ ranges from 0.9 to 1.1 with uniform distribution. Market price d is equal to 1 dollar per kilo records.

1) PERFORMANCE OF THE STACKELBERG GAME

In this paper, we present Stackelberg game approach to formulate the interaction between the data miner and local users in data trading. More specifically, in the Stackelberg game based incentive mechanism, the data miner acts as a leader to reward access privileges authorized by the users while the users are followers to determine the amounts of data records granted to the data miner accordingly. To reach a unique Stackelberg equilibrium, we propose primal-dual Lagrangian method for the data miner to determine reward policy in the game.

We first evaluate the related method performance. We repeat the algorithm 500 times and find that facing with 100 users, the method executed by the data miner can confirm the whole reward policy for all the users finally. At this time, both the data miner and local users have acquired their best responses, namely, reaching the Stackelberg equilibrium. We randomly select 10 users for the observation and compare the reward parameters in problem (6) solved by typical CVX tool and our method, as shown in Fig. 4. Clearly, there are few differences among the reward parameters calculated by various methods. This demonstrates the effectiveness of our primal-dual Lagrangian method for the solution of problem (6).

Besides, we compare performances between the Stackelberg game based scheme and maximum-amount purchasing (MAP) scheme in data trading, as shown in Fig. 5. In the MAP scheme, the data miner is suggested to access to their whole data records of local users, namely, making $x_i^* = \min(X_i, y_{\max})$. In practice, the scheme is convenient to fully satisfy the crucial data collection requirement denoted by $Y_{\min}$. However, the simple scheme neglects the revenue for the data miner in data trading. For optimization, the goal in the problem is related with overall revenue of the data miner. The revenue is increased when the data miner is of higher capability to transform data values into economic assets after performing the data mining task.

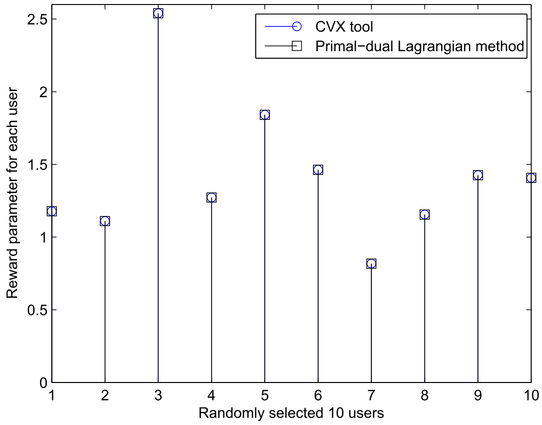


FIGURE 4. Method validation regarding to primal-dual Lagrangian method for reaching the Stackelberg equilibrium.

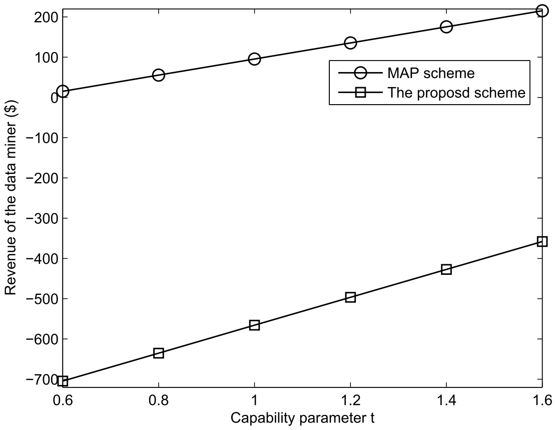
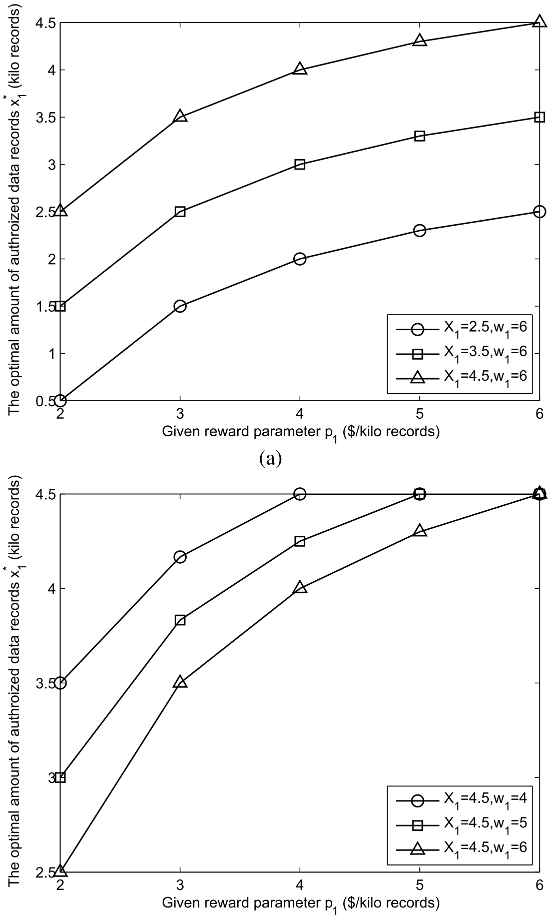


FIGURE 5. Comparison of revenue of the data owner with respect to different schemes.

Differently to the MAP scheme, our scheme optimizes data collection in terms of adjusting reward policy for accessing to enough data records and maximizing the revenue simultaneously. As a consequence, the revenue is greatly superior in our proposed scheme and dynamically improved with the increasing capability parameter t. For example, when $t = 1$, the revenue is negative in the MAP scheme while that is 95 dollar in our proposed scheme. This means with the considerate optimization, our scheme outperforms the baseline scheme and shows great advantages in maximizing the revenue for the data miner in data trading.

2) IMPACTS TO THE STACKELBERG GAME

Here, we study the impacts of different parameters for the best response of a user (follower) in the Stackelberg game based



(a)

(b)

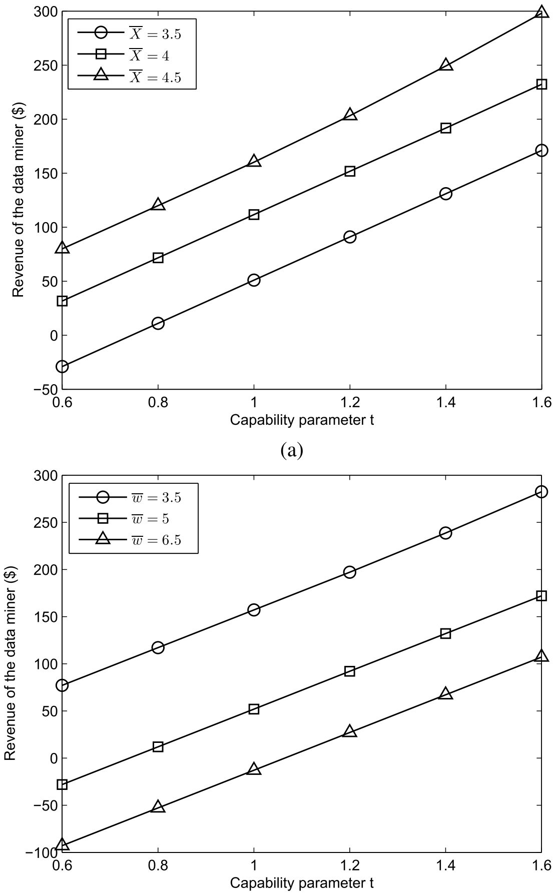
FIGURE 6. Performance comparison of the best response of a follower with respect to different p_1, X_1 and w_1. (a) Different values of the amount of available data records X_1. (b)Different values of individual willingness about data trading w_1.

scheme, which are illustrated by Fig. 6. We choose the first user for observations. Based on Eqn. (5), the best response of the user when determining the optimal amount of authorized data records to the data miner, x_1^*, is influenced the following parameters: given reward parameter p_1, the total amount of available data records X_1 and the individual willingness for data protection rather than data trading w_1.

It is obvious that with the increasing value of the reward parameter, the user would like to allow more data records to be published for earing more economic benefits. So x_1^* is increased to improve the utility with the higher value of p_1. Meanwhile, when having more data records, the user can acquire improved privacy gains in case of publishing the same amount of data records. In this way, the user also tends to permit more data records granted to the data miner because of less privacy losses. As shown in Fig. 6(a), with the same reward parameter: $p_1 = 4$ and $w_1 = 6$, the optimal amount of authorized data records x_1^* is improved with about 100% when the value of X_1 is changed from 2.5 to 4.5 kilo records.

As for the willingness to indicate the level of data protection, the parameter plays a negative impact for the faclitation of data trading. The higher value of w_1 means that the user pays more attention to prevent healthcare data well from being accessed for avoiding potential privacy threats in the utility function. As w_1 is increased from 4 to 6, the user prefers to reduce the amount of authorized data records under the condition $p_1 = 3$ and $X_1 = 4.5$. The decline in percentages regarding to the amount is more than 16% in Fig. 6(b).

In turn, the best response of the data miner (leader) corresponding to the final revenue is influenced by the internal and external parameters. The internal parameter is mainly represented by capability parameter t, to express the earning power in performing the data mining task. The external parameters are related with status information of all the users in data trading. Here, we use the average values of X and w, namely, $\overline{X}$ and $\overline{w}$, to mainly indicate the status information of local users in data trading.



(a)

(b)

FIGURE 7. Performance comparison of the best response of the leader with respect to different t, $\overline{X}$ and $\overline{w}$. (a) Different average values of the amount of available data records belonging to local users. (b) Different average values of individual willingness of each user uninteresting in data trading.

As shown in Fig. 7, we study the impacts of different parameters for the best response of the data miner in the scheme. On one hand, the data miner earns more when the capability parameter is strengthened, namely, the value of t is increased. In addition, user i prefers to publish individual healthcare data to the data miner in data trading when the total amount of available data records X_i is sufficient enough. With the same reward policy, the data miner can access to more data records as the average value of X is significantly higher. This leads to greater revenue for the data miner under the given condition of t. We take that $\overline{X}$ is improved from 3.5 to 4 kilo records as an example. In Fig. 7(a), when t is 1, the increment in percentages of the revenue is about 117% at this time.

Besides, the revenue is also influenced by the increasing unwillingness of each user for data trading. If the value of $\overline{w}$ is high, most of the users are uninterested in allowing the data miner access to their healthcare data via data trading. To meet the data collection requirement, the data miner should offer more incentive. As a result, the revenue for the data miner is lessen sharply accordingly. For example, under the condition of $t = 1.2$, the revenue is decreased from 92 dollar to 27 dollar as the value of $\overline{w}$ is increased from 5 to 6.5.

In summary, the above numerical results demonstrate that the Stackelberg game approach is feasible and effective for both the data miner and users in data trading.

VII. CONCLUSION

In this paper, we leverage edge computing to present a secure and efficient system, called by EdgeCare, for mobile healthcare systems. Thus, decentralized and collaborative data management is achieved. In EdgeCare, specialized LAs act as trusted authorities to schedules edge server for processing healthcare data with security guarantee and facilitating feasible data trading. To enable EdgeCare, the hierarchical architecture of EdgeCare is proposed and discussed to ensure the feasibility of secure and efficient data management in the decentralized mobile environment. The EMR management scheme and the corresponding protocol workflows are also elaborately designed to handle healthcare data processing with security and privacy requirements. Moreover, we study the optimization problem about decentralized data trading in EdgeCare. The interaction between a data collector and local users is formulated by Stackelberg game to closely approach the real data trading environment. Finally, numerical results with security analysis are provided to demonstrate that EdgeCare has significant advantages in security protection for healthcare data, and support efficient data trading.

REFERENCES

[1] O. S. Kemkarl and D. P. B. Dahikar, "Can electronic medical record systems transform health care? Potential health benefits, savings, and cost using latest advancements in ICT for better interactive healthcare learning," *Int. J. Comput. Sci. Commun. Netw.*, vol. 2, nos. 3–6, pp. 453–455, 2012.

[2] M. M. Baig, H. GholamHosseini, and M. J. Connolly, "Mobile healthcare applications: System design review, critical issues and challenges," *Australas. Phys. Eng. Sci. Med.*, vol. 38, pp. 23–38, Mar. 2015.

[3] P. Hunter, "The big health data sale: As the trade of personal health and medical data expands, it becomes necessary to improve legal frameworks for protecting patient anonymity, handling consent and ensuring the quality of data," *EMBO Rep.*, p. e201642917, Nov. 2016.

[4] S. Armstrong, "Data, data everywhere: The challenges of personalised medicine," *BMJ*, vol. 359, p. j4546, Oct. 2017.

[5] L. A. Tawalbeh, R. Mehmood, E. Benkhlifa, and H. Song, "Mobile cloud computing model and big data analysis for healthcare applications," *IEEE Access*, vol. 4, pp. 6171–6180, 2016.

[6] H. Kaur, M. A. Alam, R. Jameel, A. K. Mourya, and V. Chang, "A proposed solution and future direction for blockchain-based heterogeneous medicare data in cloud environment," *J. Med. Syst.*, vol. 42, p. 156, Aug. 2018.

[7] L. Zhu, Y. Wu, K. Gai, and K.-K. R. Choo, "Controllable and trustworthy blockchain-based cloud data management," *Future Gener. Comput. Syst.*, vol. 91, pp. 527–535, Feb. 2019.

[8] M. N. K. Boulos, J. T. Wilson, and K. A. Clauson, "Geospatial blockchain: Promises, challenges, and scenarios in health and healthcare," *Int. J. Health Geogr.*, vol. 17, p. 25, Jul. 2018.

[9] Q. Xia, E. B. Sifah, S. Smahi, S. Amofa, and X. Zhang, "BBDS: Blockchain-based data sharing for electronic medical records in cloud environments," *Information*, vol. 8, no. 2, p. 44, 2017.

[10] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using blockchain for medical data access and permission management," in *Proc. 2nd Int. Conf. Open Big Data (OBD)*, Aug. 2016, pp. 25–30.

[11] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards," *ETSI White Paper*, vol. 11, no. 11, pp. 1–16, 2015.

[12] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.

[13] X. Huang, R. Yu, J. Kang, and Y. Zhang, "Distributed reputation management for secure and efficient vehicular edge computing and networks," *IEEE Access*, vol. 5, pp. 25408–25420, 2017.

[14] N. Kumar, S. Zeadally, and J. J. P. C. Rodrigues, "Vehicular delay-tolerant networks for smart grid data management using mobile edge computing," *IEEE Commun. Mag.*, vol. 54, no. 10, pp. 60–66, Oct. 2016.

[15] W. Shi and S. Dustdar, "The promise of edge computing," *Computer*, vol. 49, no. 5, pp. 78–81, 2016.

[16] J. Kang, X. Huang, R. Yu, Y. Zhang, and S. Gjessing, "Hierarchical mobile cloud with social grouping for secure pervasive healthcare," in *Proc. 17th Int. Conf. E-Health Netw., Appl. Services (HealthCom)*, Oct. 2015, pp. 609–614.

[17] D. Thilakanathan, S. Chen, S. Nepal, R. A. Calvo, and L. Alem, "A platform for secure monitoring and sharing of generic health data in the cloud," *Future Generat. Comput. Syst.*, vol. 35, pp. 102–113, Jun. 2014.

[18] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Syst. J.*, vol. 11, no. 1, pp. 88–95, Mar. 2017.

[19] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 1, pp. 131–143, Jan. 2013.

[20] T. Nugent, D. Upton, and M. Cimpoesu, "Improving data transparency in clinical trials using blockchain smart contracts," *Research*, vol. 5, p. 2541, Oct. 2016.

[21] A. Abbas and S. U. Khan, "A review on the state-of-the-art privacy-preserving approaches in the e-health clouds," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1431–1441, Apr. 2014.

[22] M. Al Ameen, J. Liu, and K. Kwak, "Security and privacy issues in wireless sensor networks for healthcare applications," *J. Med. Syst.*, vol. 36, pp. 93–101, Feb. 2012.

[23] M. A. Alkureishi *et al.*, "Impact of electronic medical record use on the patient-doctor relationship and communication: A systematic review," *J. Gen. Internal Med.*, vol. 31, pp. 548–560, May 2016.

[24] G. Manogaran, C. Thota, D. Lopez, and R. Sundarasekar, "Big data security intelligence for healthcare industry 4.0," in *Cybersecurity for Industry 4.0: Analysis for Design and Manufacturing*. Cham, Switzerland: Springer, 2017, pp. 103–126.

[25] C. C. Aggarwal and P. S. Yu, "A general survey of privacy-preserving data mining models and algorithms," in *Privacy-Preserving Data Mining*. Boston, MA, USA: Springer, 2008, pp. 11–52.

[26] L. Xu, C. Jiang, Y. Chen, J. Wang, and Y. Ren, "A framework for categorizing and applying privacy-preservation techniques in big data mining," *Computer*, vol. 49, no. 2, pp. 54–62, Feb. 2016.

[27] R. H. Khokhar, B. C. Fung, F. Iqbal, D. Alhadidi, and J. Bentahar, "Privacy-preserving data mashup model for trading person-specific information," *Electron. Commerce Res. Appl.*, vol. 17, pp. 19–37, May/Jun. 2016.

[28] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," *IEEE Access*, vol. 2, pp. 1149–1176, Oct. 2014.

[29] S. Romanosky and A. Acquisti, "Privacy costs and personal data protection: Economic and legal perspectives," *Berkeley Technol. Law J.*, vol. 24, no. 1, p. 1061, 2009.

[30] L. Xu, C. Jiang, Y. Chen, Y. Ren, and K. J. R. Liu, "Privacy or utility in data collection? A contract theoretic approach," *J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1256–1269, Oct. 2015.

[31] S. Spiekermann, A. Acquisti, R. Böhme, and K.-L. Hui, "The challenges of personal data markets and privacy," *Electron. Markets*, vol. 25, no. 2, pp. 161–167, 2015.

[32] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97–107, Jan. 2014.

[33] J. B. Cruz, Jr., "Survey of nash and stackelberg equilibrim strategies in dynamic games," in *Ann. Econ. Social Meas.*, vol. 4, no. 2, pp. 339–344, 1975.

[34] M. T. Siponen and H. Oinas-Kukkonen, "A review of information security issues and respective research contributions," *SIGMIS Database*, vol. 38, pp. 60–80, Feb. 2007.

Authors' photographs and biographies not available at the time of publication.